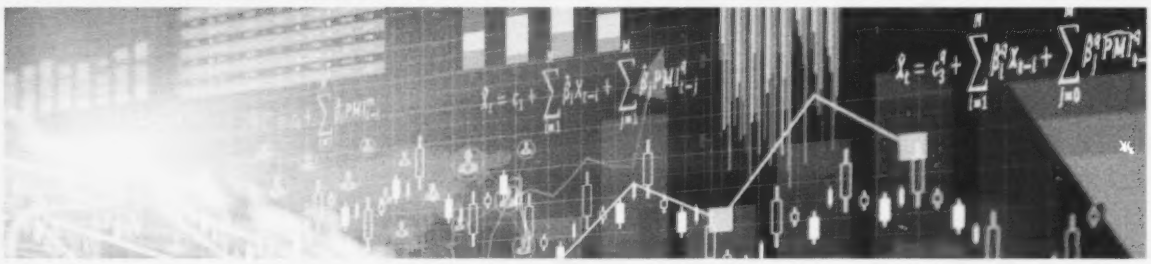
BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel 2018-25

On the Evolution of the United Kingdom Price Distributions



by Ba M. Chu, Kim P. Huynh, David T. Jacho-Chávez and Oleksiy Kryvtsov

Bank of Canada staff working papers provide a forum for staff to publish work-in-progress research independently from the Bank's Governing Council. This research may support or challenge prevailing policy orthodoxy. Therefore, the views expressed in this paper are solely those of the authors and may differ from official Bank of Canada views. No responsibility for them should be attributed to the Bank.

www.bank-banque-canada.ca

Bank of Canada Staff Working Paper 2018-25

June 2018

On the Evolution of the United Kingdom Price Distributions

by

Ba M. Chu,[1] Kim P. Huynh,[2] David T. Jacho-Chávez[3] and Oleksiy Kryvtsov[4]

[1] Department of Economics
Carleton University
Ottawa, Ontario, Canada K1S 5B6
ba_chu@carleton.ca

[2] Currency Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
khuynh@bankofcanada.ca

[3] Department of Economics
Emory University
Atlanta, Georgia, USA 30322
djachocha@emory.edu

[4] International Economic Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
okryvtsov@bankofcanada.ca

ISSN 1701-9397

© 2018 Bank of Canada

Acknowledgements

We thank participants of various seminars and workshops for helpful comments. We are grateful to the UK Office for National Statistics, especially Derek Bird and Ainslie Restieaux for their assistance with this project; Rosie Foster and Rowan Kelsoe for their detailed responses to our queries regarding the data. Boyan Bejanov and Claudiu Motoc provided excellent scientific computation advice on using SQLite to handle large datasets and using npRmpi and snow parallel procedures on the EDITH high performance cluster. Tony Chernis, Geoff Halmo, Anderson Nzabandora, and Shane Wood provided excellent research assistance. The views expressed in this paper are those of the authors. No responsibility for them should be attributed to the Bank of Canada. All remaining errors are the responsibility of the authors.

Abstract

We propose a functional principal components method that accounts for stratified random sample weighting and time dependence in the observations to understand the evolution of distributions of monthly micro-level consumer prices for the United Kingdom (UK). We apply the method to publicly available monthly data on individual-good prices collected in retail stores by the UK Office for National Statistics for the construction of the UK Consumer Price Index from March 1996 to September 2015. In addition, we conduct Monte Carlo simulations to demonstrate the effectiveness of our methodology. Our method allows us to visualize the dynamics of the price distribution and uncovers interesting patterns during the sample period. Further, we demonstrate the efficacy of our methodology with an out-of-sample forecasting algorithm that exploits the time dependence of distributions. Our out-of-sample forecast compares favorably with the random walk forecast.

Bank topics: Econometric and statistical methods; Inflation and prices
JEL codes: C14, C83, E31, E37

Résumé

Nous proposons une méthode d'analyse en composantes principales fonctionnelles qui rend compte de la pondération de l'échantillonnage aléatoire stratifié et de la dépendance temporelle des observations dans le but de comprendre l'évolution des distributions de données microéconomiques mensuelles sur les prix à la consommation au Royaume-Uni. Nous appliquons la méthode à des données mensuelles sur les prix par produit recueillies par l'Office for National Statistics dans les commerces de détail entre mars 1996 et septembre 2015 afin de construire l'indice des prix à la consommation du Royaume-Uni. De plus, nous effectuons des simulations de Monte-Carlo pour démontrer l'efficacité de notre méthodologie. Notre méthode nous permet de visualiser la dynamique de la distribution de prix et met en lumière des phénomènes intéressants pendant la période considérée. Par ailleurs, nous démontrons l'efficacité de notre méthodologie à l'aide d'un algorithme de prévision hors échantillon qui exploite la dépendance temporelle des distributions. Notre prévision hors échantillon se compare favorablement à la prévision issue d'une marche aléatoire.

Sujets : Méthodes économétriques et statistiques; Inflation et prix
Codes JEL : C14, C83, E31, E37

Non-technical Summary

Motivation and Question

This paper proposes a method to visualize the dynamics of the distribution of the United Kingdom (UK) Consumer Price Index (CPI). This method allows one to decompose the distribution into a cross-section and time-series. We apply this methodology to over 26 million observations from the UK CPI for the period March 1996 to September 2015.

Methodology

This paper proposes a functional principal components analysis (FPCA) to decompose distributions into cross-section and time-series components. In addition, we allow for stratified random sample weighting to account for sample weights that accompany CPI data. We also demonstrate how to efficiently compute this FPCA decomposition with the usage of open source statistical software R using a high performance cluster (HPC).

Key Results and Contributions

In the first application, we explore how the functional principal components comove with the unemployment and inflation rates—two key macroeconomic variables. We find that the dynamics of the higher order moments are as important as the first moment in driving temporal changes in price and price-change distributions. For example, price dispersion decreases and the tails thin out in the wake of the 2009 recession, but these effects are reversed over the ensuing 2011 recession. At the same time price changes become more dispersed, suggesting that the time of economic turmoil triggers across-the-board price adjustments.

In the second application, we demonstrate the efficacy of our methodology by applying it to an out-of-sample density forecasting exercise for the price-level and price-change distributions. We conduct this exercise for one- and three-month forecast horizons as they are the standard time horizons between the release of official price data. We find that our FPCA model-based forecasts compare favorably with the benchmark random-walk or the 'no-change' forecast.

1 Introduction

This paper investigates the evolution of the underlying distribution of price quotes collected to construct the Consumer Price Index (CPI) for the United Kingdom. In all countries, the CPI is the most commonly used measure of the average price that households pay for their consumption, and it is also widely used as a measure of the changes in the cost of living or inflation. The ubiquity of the CPI in the everyday lives of consumers and firms makes it one of the most watched, reported, and discussed statistic in a modern economy. Statistical agencies such as the Bureau of Labor Statistics (BLS) in the United States, the Office for National Statistics (ONS) in the United Kingdom, Statistics Canada, and Eurostat devote considerable resources to the efficient collection and dissemination of these data. Central banks all over the world rely on the CPI to analyze and formulate economic policies, especially countries with an explicit inflation target such as the United States (US), United Kingdom (UK), and Canada.

The population of prices underlying the CPI exhibits much more complex dynamics than can be captured by an aggregated index. These dynamics manifest themselves along two key dimensions: (1) prices do not move in a synchronous manner across goods and services: for example, food and energy prices tend to be more volatile than other prices; and (2) relative expenditures for goods and services are also time-varying, as households shift their consumption in response to price changes, discounts, seasons and holidays, stockouts, and changes in quality over the business cycle (see Klenow and Malin, 2010 and Nakamura and Steinsson, 2013 for comprehensive reviews of microeconomic evidence on price dynamics).

While in practice different measures of the CPI have been advocated for capturing some of these dynamics, such measures shed little light on the evolution of price distributions that underlie the CPI (see Diewert, 2012 for a detailed discussion of various CPI measures in the UK). For example, little is known about which part of the distribution drives the changes in the means or whether higher order moments such as skewness or kurtosis play a role in changes in distributions. For example, early empirical work using evidence for sector-level price indexes and for micro prices for selected products focused on documenting the extent of price or price change dispersion, including contributions by Lach and Tsiddon (1992), Ball and Mankiw (1995), and Debelle and Lamont (1997). Recent studies of the moments of price and price change distributions focused on price dispersion (Peterson and Shi (2004), Kaplan and Menzio (2015), and Sheremirov (2016)), price-change dispersion (Vavra, 2013 and Berger and Vavra, 2018), skewness of price distribution (Chen et al., 2008 and Sheremirov, 2016), and kurtosis of price changes (Midrigan, 2011 and Alvarez et al., 2016).

Understanding these complex movements of relative prices and expenditure weights requires information about micro-level prices and their respective weights in the CPI. A limited availability of such data, however, hinders the advancement of tools and methods that could be used by economists who study price dynamics. This paper utilizes monthly data on individual-good prices collected in retail stores by the ONS for the construction of the UK CPI from February 1996 to September 2015. These are the first publicly available

monthly data on individual-good prices collected by a national statistical agency, and can be downloaded directly from the ONS website: http://www.ons.gov.uk/ons/datasets-and-tables/index.html. We utilize this dataset to estimate the evolution of monthly consumer price distributions for the UK.

The main challenge for this estimation stems from the complexity of the relative price and weight movements. We tackle this challenge by developing functional principal component analysis (FPCA) with complex survey design methodology. Let $f_1, \ldots, f_T$ denote the super-population distributions of price (or price changes) in the UK that can be recovered from this micro-level consumer price data. We build over Kneip and Utikal's (2001) model. This model uses the well-known Karhunen–Loève decomposition to represent each super-population density at time t, f_t, in terms of the principal functional components:

$$f_t = f_\mu + \sum_{j=1}^{J} \theta_{t,j} g_j. \tag{1}$$

Here the mean density, $f_\mu = T^{-1} \sum_{t=1}^{T} f_t$, captures the underlying common intertemporal distribution, while the J dynamic strength coefficients, $\theta_{t,j}$, capture the dynamics of the underlying distributions, and the common functional components, or basis functions, g_j, provide information about the part of the distribution that changes. FPCA tracks how a density evolves over time by providing a dynamic structure as opposed to standard static visualization of estimated cross-section distributions.

Although Kneip and Utikal (2001) and Huynh et al. (2011) already proposed a way to estimate the model in (1) with repeated independent samples, the problem we consider here involves adapting the original methodology to complex survey data that are longitudinal in nature, i.e., recorded prices from the same strata that are potentially correlated over time. These data features require a modification of Kneip and Utikal's (2001) original estimators to include survey weights so that the dynamics of the super-population densities are described by FPCA, and not the dynamics of the densities induced by the sampling scheme. We also introduce an additional bias correction to account for temporal dependence, which extends Kneip and Utikal's (2001) original results to handle temporally correlated samples. The paper demonstrates the consistency of the resulting estimators of the score coefficients and the pointwise asymptotic normality of the functional components. A data-driven procedure is then proposed to choose the optimal number (J^*) of principal components based on a density forecast exercise.

This methodology allows for a simple representation and visualization of the temporal dynamics of price distributions. According to (1), the distributions of the detrended prices or the month-to-month price changes can be decomposed into functional principal components. The time-variation of the corresponding $\theta_{t,j}$ can be used to visualize the evolution of the principal components over time. We find that the first three components account for about 10 and 8% of the variation in the price and price-change distributions in the UK CPI data, respectively. The dynamics of the first three components, as described by $\theta_{t,j}$, are correlated with the key macroeconomic variables, and also autocorrelated.

In the second part of the paper, we explore these properties of the functional principal components in two applications that study the comovement of the UK price distributions with macroeconomic variables and forecasting of the UK consumer prices.

In the first application, we explore how the functional principal components comove with the unemployment and inflation rates—two key macroeconomic variables. Here, an advantage of FPCA naturally stems from its decomposition of price (or price-change) distributions into the time-varying dynamic strength coefficients and time-invariant common functional components, given by $\theta_{t,j}$ and g_j respectively. Namely, the first deformation, $\theta_{t,1}g_1$, is closely related to the evolution of the first moment of the distribution, and higher order components, $\{\theta_{t,j}g_j\}_{j>1}$, summarize contributions of higher order moments, such as the dispersion, skewness, and kurtosis of the distribution. We find that the dynamics of the higher order moments are as important as the first moment in driving temporal changes in price and price-change distributions. For example, price dispersion decreases and the tails thin out in the wake of the 2009 recession, but these effects are reversed over the ensuing 2011 recession. At the same time price changes become more dispersed, suggesting that the time of economic turmoil triggers across-the-board price adjustments. In all, the facts unraveled with the help of FPCA can be influential for sorting out models of inflation and business cycles used in academia and central banks, because their predictions are inherently linked to assumptions about individual price adjustments in response to economic shocks.

In the second application, we demonstrate the efficacy of our methodology by applying it to an out-of-sample density forecasting exercise for the price-level and price-change distributions. We conduct this exercise for one- and three-month forecast horizons as they are the standard time horizons between the release of official price data. We find that our FPCA model-based forecasts compare favorably with the benchmark random-walk or the 'no-change' forecast.

Section 2 provides a description of the UK price data used in the paper. Section 3 discusses the FPCA with complex survey design methodology. Section 4 conducts the analysis of the distribution dynamics and provides numerical evidence of the proposed estimators in a simulation study. Section 5 discusses the results of the simulations and studies the comovement of the UK price distributions with macroeconomic variables. This section also suggests how to perform model validation through a density forecasting exercise and offers some sensitivity analysis. Finally, Section 6 concludes.

2 UK Consumer Price Index Data

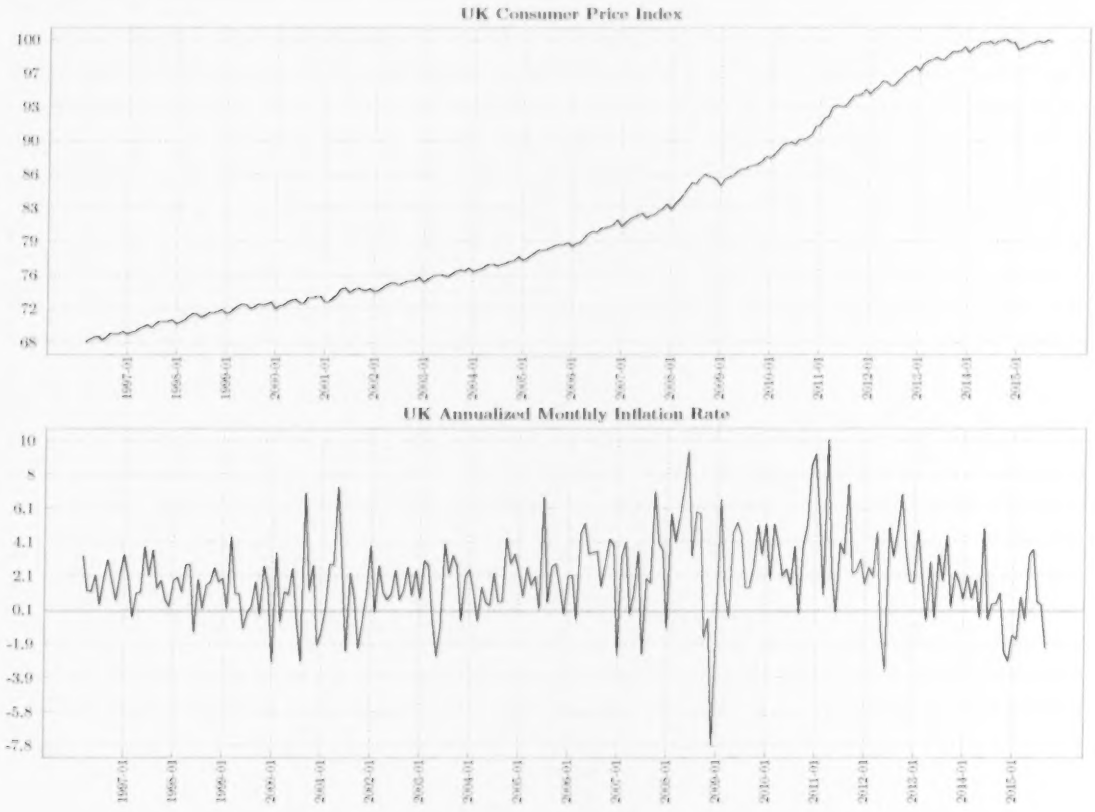
To construct the CPI, the ONS surveys prices for goods and services that are included in the household final monetary consumption expenditure component of the UK National Accounts. The survey excludes the housing portion of consumer prices, such as mortgage interest payments, house depreciation, insurance, and other house purchase fees. A detailed description of the data underlying the CPI, the statistical methodology used, collection and validation of prices, and calculation of weights can be found in the

ONS (2014) or Clews et al. (2014). In total, the survey includes prices for over 1,100 representative items (e.g., onions, men's suits, single beds), which are collected locally or centrally, for more than 14,000 retail stores across 12 geographical regions (London, South East, South West, Eastern, East Midlands, West Midlands, Yorkshire and the Humber, North West, North East, Scotland, Wales, and Northern Ireland). Each item is assigned a weight that reflects its relative importance in households' consumption expenditures. Weights are calculated based on the Household Final Monetary Consumption Expenditure and ONS Living Cost and Food Survey. Changes in expenditure weights over time reflect gradual shifts in expenditure composition of households' consumption baskets. The data published on the ONS website include only locally collected prices, covering about 57% of the UK CPI basket.

Observations are stratified by shop type, by region, and by region and shop type. Two types of shops are identified: retailers with fewer than 10 outlets are classified as independents, while retailers with 10 or more outlets are classified as multiples. This study utilizes only the samples stratified by region and shop type yielding a total of 24 strata (L) to better capture price heterogeneity in the population. The sample period includes 235 months (T), from March 1996 till September 2015. The total number of observations is over 26 million, or about 110,000 per month.

Figure 1 displays the UK CPI and annualized monthly inflation rate. The light yellow areas highlight the UK's late 2000s (2008–04/2009–09) and the double-dip (2011–10/2012–06) recessions. The red solid line represents zero. The figure shows that the CPI level in the UK displays significant trend and cyclical variations across time. The price level grew by an average of 2.0% per year between 1996 and 2015, and the standard deviation of their changes was 0.4 percentage points per month; in particular, inflation fell by 2 to 3 percentage points in annualized terms during each of the two recessions in the 2000s. These changes are due to a combined effect of micro-level time co-variations of prices across different products, stores, and locations. As we will show below, much of the information on such complex micro-level price co-variations is contained in the dynamics of the distribution of prices.

Figure 1: UK Monthly CPI and Inflation. The top graph plots the monthly CPI (source: http://www.ons.gov.uk), and the bottom graph plots the implied annualized seasonally adjusted monthly inflation in percentages. The base year is 2015. The light yellow areas highlight the UK's late 2000s (2008–04/2009–09), and the double-dip (2011–10/2012–06) recessions.



To study these dynamics, we construct the following variables. Let t denote a month, k denote an item-stratum bin, and i denote an individual product, so that $\breve{p}_{k,i}(t)$ represents the natural logarithm of the price of a product or service i (in nominal British pounds) in month t, for item-stratum bin k; and $\Delta\breve{p}_{k,i}(t) = \breve{p}_{k,i}(t) - \breve{p}_{k,i}(t-1)$ is the change of the natural logarithm of price for that product from month $t-1$ to t. We also define the detrended price levels, $\breve{p}^*_{k,i}(t) = \breve{p}_{k,i}(t) - \overline{\breve{p}}_k(t)$, where $\overline{\breve{p}}_k(t)$ is the across-time linear trend of $\breve{p}_{k,i}(t)$ estimated for each item-stratum bin k. As a point of comparison, Figure 1 in the supplemental material (Chu et al., 2018) displays estimates of the population mean, median and inter-quartile range of $\breve{p}_{k,i}(t)$ for each of the 235 months.

Considerable variations of prices at store and product levels suggest that the observed dynamics of distributions of $\breve{p}_{k,i}(t)$ or $\Delta\breve{p}_{k,i}(t)$ could be driven by a subset of stores or goods with particularly volatile prices, creating numerical difficulties when recovering the

overall movement of the super-population distribution of prices. Furthermore, in a given month roughly 84% of prices do not change from the previous month. To address these issues, we proceed by standardizing price levels and price changes as suggested in Klenow and Kryvtsov (2008) and defined as follows:

$$x^*_{k,i}(t) = (\ddot{p}^*_{k,i}(t) - \overline{\ddot{p}^*}_k)/\widehat{\sigma}_{\ddot{p}^*_k} \text{ and } \Delta x_{k,i}(t) = (\Delta\ddot{p}_{k,i}(t) - \overline{\Delta\ddot{p}}_k)/\widehat{\sigma}_{\Delta\ddot{p}_{k,i}}, \tag{2}$$

where $\overline{\ddot{p}^*}_k$ and $\overline{\Delta\ddot{p}}_k$ are across-time sample means of $\ddot{p}^*_{k,i}(t)$ and non-zero $\Delta\ddot{p}_{k,i}(t)$ computed for each item-stratum k, and $\widehat{\sigma}_{\ddot{p}^*_{k,i}}$ and $\widehat{\sigma}_{\Delta\ddot{p}_{k,i}}$ for the standard deviations. Hereafter, we refer to $x^*_{k,i}(t)$ and $\Delta x_{k,i}(t)$ as the 'log(Price)' and 'Δlog(Price)' respectively. Their underlying super-population distributions in each month, generally denoted by f_t, are the focus of our analysis.

3 Functional Principal Component Analysis (FPCA) of Densities

Kneip and Utikal (2001) provide a method to identify $\{g_j\}_{j=1}^J$ and $\{\{\theta_{t,j}\}_{j=1}^J\}_{t=1}^T$ on the right-hand side of (1). Let $\mathcal{M}$ be the variance-covariance matrix with the elements:

$$\mathcal{M}_{t,s} = \langle f_t - f_\mu, f_s - f_\mu \rangle = M_{t,s} - \frac{1}{T}\sum_{i=1}^{T}(M_{t,i} + M_{i,s}) + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} M_{t,s}, \tag{3}$$

where $M_{t,s} = \langle f_t, f_s \rangle = \int_{\mathcal{X}} f_t(x) f_s(x) w(x) dx$, $w(x) > 0$ is some continuous, uniformly bounded weighting function $\forall x \in \mathcal{X}$, and $\mathcal{X}$ represents a compact support. Kneip and Utikal (2001) show that $\theta_{t,j}$ and g_j can be written as:

$$\theta_{t,j} = \lambda_j^{1/2} p_{t,j},\ g_j = \frac{\sum_{t=1}^{T} \theta_{t,j} f_t}{\sum_{t=1}^{T} \theta_{t,j}^2}, \tag{4}$$

where $p_1 = (p_{1,1}, \ldots, p_{T,1})^\top$; ...; and $p_j = (p_{1,J}, \ldots, p_{T,J})^\top$ are the eigenvectors of $\mathcal{M}$; and $\lambda_1 \geq \lambda_2 \geq \cdots$ are the eigenvalues of $\mathcal{M}$. Thus, in view of (1) and (4) from classical PCA one has:

$$\sum_{t=1}^{T} \theta_{t,j} = 0,\ \sum_{t=1}^{T} \theta_{t,j}\theta_{t,l} = 0 \text{ if } j \neq l,\ \sum_{t=1}^{T} \theta_{t,j}^2 = \lambda_j,\ j = 1, \ldots, J, \tag{5}$$

$$\int_{\mathcal{X}} g_r(x) dx = 0. \tag{6}$$

3.1 FPCA with Complex Survey Data

Equations (3) and (4) are the basis to construct feasible estimators of each component in the right-hand side of the model (1). However, we face two data features that need to be

addressed when constructing these estimators, namely, having a stratified random sample each month, and the fact that prices are recorded from the same strata units through time. In this section we describe how Kneip and Utikal's (2001) original estimators can be adapted to these situations, and we also investigate their asymptotic behavior when accounting for these new data features.

Let $\|v\|$ be the Euclidean norm of $v = (v_1, \ldots, v_{\dim(v)})^\top$; $v^\wedge = \min(v_1, \ldots, v_{\dim(v)})$; $v^\vee = \max(v_1, \ldots, v_{\dim(v)})$; $U_{k,i}(t,x) = K_{h_k}(x - X_{k,i}(t))$; and $\langle U_{k,i}(t), U_{\ell,j}(s) \rangle = \int U_{k,i}(t,x) U_{\ell,j}(s,x) w(x) dx$, where $w(x)$ is the norm weight and $K_h(\cdot) = h^{-1} K(\cdot/h)$ is some kernel function defined below; h_k denotes a smoothing parameter; $K(\cdot)$ satisfies Assumption 2 below; $K^*(z) = \int K(y) K(z + y)\, dy$; $O_p(\cdot)$, $O_{a.s.}(\cdot)$, $o_p(\cdot)$, and $o_{a.s.}(\cdot)$ are symbols for stochastic orders of magnitude ('in probability' and 'almost surely' respectively) taken with respect to the distributional law of the actual data; $\xrightarrow{d}$ denotes convergence in distribution (under the distributional law of the actual data); *w.p.*1 stands for 'with probability (w.r.t. the survey sampling scheme) approaching 1.' Finally, all the expectations are taken with respect to the distributional law of the actual data, so that we just write $E[\cdot] \equiv E_X[\cdot]$ to simplify notation.

3.1.1 Stratified Random Sampling, Dependence, and Estimation

Further notation is now introduced in order to describe the estimators and their asymptotic properties. Let $N_1, \ldots, N_L$ denote the sizes of L sub-populations, $U_1(t), \ldots, U_L(t)$, available at time t, and let $N = \sum_{k=1}^{L} N_k$ be the size of the entire super-population. Here we make the assumption that the population sizes are invariant over time so as to keep all calculations involving multiple summations tractable. This assumption does not invalidate the theoretical results obtained in this paper, and can be relaxed at the expense of more cumbersome notations and algebra.

We assume that the elements in each $U_k(t)$ are realizations of the i.i.d. random variables, $X_{k,1}(t), \ldots, X_{k,N_k}(t)$, distributed according to a density, $f_{t,k}$; also, $X_{k,i}(t)$ and $X_{\ell,j}(t)$, where $i = 1, \ldots, N_k$ and $j = 1, \ldots, N_\ell$, are pairwise independent for all $k \neq \ell$. The stochastic process $X_{k,i}(t)$ is strongly mixing in the sense of Rosenblatt (1956) for each $k \in [1, L]$ and $i \in [1, N_k]$; this process is generally non-stationary as its density, $f_{t,k}$, will not vary over time in case $X_{k,i}(t)$ is a strictly stationary process.

The stratum (sub-population) weights are defined by $W_k = N_k/N$. The super-population density, f_t, then becomes a mixture of the sub-population densities:

$$f_t(x) = \sum_{k=1}^{L} W_k f_{t,k}(x). \tag{7}$$

It is assumed that random samples of n_k observations are taken *without replacement* from each stratum, $U_k(t)$, denoted by $S_k(t)$; then $S(t) = \bigcup_{k=1}^{L} S_k(t)$ is a stratified random sample of size $n = \sum_{k=1}^{L} n_k$ from the super-population, $U(t) = \bigcup_{k=1}^{L} U_k(t)$. This survey sampling design can be represented by random indicators, $I_{k,i}(t) = 1$ if unit i of the stratum k belongs to S_k at time t, and zero otherwise. Because a part of the super-

population can be sampled, Buskirk (1998, 1999), Bellhouse and Stafford (1999), and Buskirk and Lohr (2005) propose to estimate $f_t(x)$ by a sample weighted kernel density (SWKD):

$$\widehat{f}_{n,t}(x) = \frac{1}{N}\sum_{k=1}^{L}\frac{N_k}{n_k}\sum_{i=1}^{N_k} I_{k,i}(t)K_{h_k}\left(x - X_{k,i}(t)\right) = \sum_{k=1}^{L} W_k \widehat{f}_{t,k}(x), \tag{8}$$

where the strata-specific bandwidth, h_k, goes to zero as the sample size becomes large, and $\widehat{f}_{t,k}(x) = n_k^{-1}\sum_{i=1}^{N_k} I_{k,i}(t)K_{h_k}\left(x - X_{k,i}(t)\right)$. Note that a single overall bandwidth h, i.e., $h_1 = \cdots, h_L = h$, can be used instead (as in our simulation study and empirical application) but this flexibility might be needed in other settings where sparseness and small samples are practical problems. Below, we assume that $n_k = O(N_k)$ for $k = 1, \ldots, L$.

The SWKD estimator in (8) is the building block of the proposed estimators of $\theta_{t,r}$ and g_r for $r = 1, 2, \ldots,$. It is used to estimate $\mathcal{M}$ by a survey sample variance-covariance matrix, i.e., a natural candidate seems to be $\widetilde{\mathcal{M}}_{t,s} = \langle \widehat{f}_{n,t} - \widehat{f}_{n,\mu}, \widehat{f}_{n,s} - \widehat{f}_{n,\mu}\rangle$ with $\widehat{f}_{n,\mu} = T^{-1}\sum_{t=1}^{T}\widehat{f}_{n,t}$. One has that:

$$\widetilde{\mathcal{M}}_{t,s} = \widetilde{M}_{t,s} - \frac{1}{T}\sum_{i=1}^{T}(\widetilde{M}_{t,i} + \widetilde{M}_{i,s}) + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{M}_{t,s}, \tag{9}$$

where $\widetilde{M}_{t,s} = \langle \widehat{f}_{n,t}, \widehat{f}_{n,s}\rangle$.

We have observed that for $t \neq s$, the candidate estimator $\widetilde{M}_{t,s}$ has some bias caused by serial correlation of $X_{k,i}(t)$. After removing this bias, we obtain a bias-corrected representation for $\widetilde{M}_{t,s}$:

$$\begin{aligned}\widehat{M}_{t,s} &= \langle \widehat{f}_{n,t}, \widehat{f}_{n,s}\rangle - \sum_{k=1}^{L} W_k^2 \frac{1}{n_k^2}\sum_{i=1}^{N_k} I_{k,i}(t)I_{k,i}(s)\langle U_{k,i}(t), U_{k,i}(s)\rangle \\ &+ \sum_{k=1}^{L} W_k^2 \frac{1}{n_k^2(n_k-1)}\sum_{i=1}^{N_k}\sum_{j\neq i}^{N_k} I_{k,i}(t)I_{k,j}(t)\langle U_{k,i}(t), U_{k,j}(s)\rangle,\end{aligned} \tag{10}$$

where the last term is needed so to ensure the same convergence rates for the bias and the asymptotic variances as in Kneip and Utikal (2001). When $t = s$, the bias-corrected representation for $\widetilde{M}_{t,t}$ is given by:

$$\begin{aligned}\widehat{M}_{t,t} &= \langle \widehat{f}_{n,t}, \widehat{f}_{n,t}\rangle - \sum_{k=1}^{L} W_k^2 \frac{1}{n_k^2}\sum_{i=1}^{N_k} I_{k,i}(t)\langle U_{k,i}(t), U_{k,i}(t)\rangle \\ &+ \sum_{k=1}^{L} W_k^2 \frac{1}{n_k^2(n_k-1)}\sum_{i=1}^{N_k}\sum_{j\neq i}^{N_k} I_{k,i}(t)I_{k,j}(t)\langle U_{k,i}(t), U_{k,j}(t)\rangle.\end{aligned} \tag{11}$$

Thus the bias-corrected estimate of $\mathcal{M}_{t,s}$ is given by:

$$\widehat{\mathcal{M}}_{t,s} = \widehat{M}_{t,s} - \frac{1}{T}\sum_{i=1}^{T}(\widehat{M}_{t,i} + \widehat{M}_{i,s}) + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widehat{M}_{t,s}. \tag{12}$$

After calculating the eigenvalues $\widehat{\lambda}_r$, and eigenvectors $\widehat{p}_r$ of $\widehat{\mathcal{M}}$ for $r = 1, \ldots, J$, one can then estimate (4) by the analogy principle:

$$\widehat{\theta}_{t,r} = \widehat{\lambda}_r^{1/2}\widehat{p}_{t,r}, \; \widehat{g}_r = \frac{\sum_{t=1}^{T}\widehat{\theta}_{t,r}\widehat{f}^*_{n,t}}{\sum_{t=1}^{T}\widehat{\theta}^2_{t,r}}, \tag{13}$$

where $\widehat{f}^*_{n,t}$ is defined as in (8) but with bandwidths $b_1, \ldots, b_L$ instead.

3.1.2 Computational Considerations

The calculation of $\widehat{\mathcal{M}}$ can be performed in an embarrassingly parallel fashion. Since $\widehat{\mathcal{M}}$ is symmetric, its $T(T-1)/2$ elements (12) require samples $S(t)$ and $S(s)$ for their calculation. Furthermore, if a new sample, i.e., $S(T+1)$, is available at time $T+1$, a new $\widehat{\mathcal{M}}$ can be readily obtained by appending a new column of $T+1$ new elements (and its transpose as a row vector) without recalculating its other elements. This is the basis of our density forecast exercise in Section 5.4.

Although the next section will provide the theoretical requirements on bandwidths, h_k and b_k, for the consistency of $\widehat{\theta}_{t,r}$ and pointwise asymptotic normality of $\widehat{g}_r$ defined in (13) to hold, we can provide a valid way to compute them in practice here. As explained earlier, in the absence of sparseness or small sample within strata, at each time period t, one can set $h_1 = \cdots = h_L = h$ and $b_1 = \cdots = b_L = b$ and then set $h = \widehat{h}^{5/4}$ and $b = \widehat{h} \times T^{-1/5}$, where $\widehat{h}$ corresponds to Duin's (1976) likelihood cross-validation bandwidth. These choices will be consistent with our theoretical developments below when the serial dependence is not very strong, and as shown in the simulation study and application they seem to work very well in practice. The supplemental material (Chu et al., 2018) provides a sensitivity analysis when $\widehat{h}$ is calculated by Silverman's (1986) rule of thumb instead.

3.2 Inference

We investigate the asymptotic behavior of the eigenvalues $\widehat{\lambda}_r$ and eigenvectors $\widehat{p}_r$ of $\widehat{\mathcal{M}}$ for $r = 1, \ldots, J$, as well as of estimators (13). To make the derivation of the asymptotics more tractable, we fix J and L, and base the asymptotics on the assumption that $N^{\wedge} = \min(N_1, \ldots, N_L)$ and $N^{\vee} = \max(N_1, \ldots, N_L)$ go to infinity, such that $N^{\wedge}/N^{\vee} = O(1)$,, $W_k = O(1)$, and $\max_{1 \le k \le L} n_k/N_k = O(1)$. We need the following standing assumptions for asymptotic analysis:

Assumption 1. *The true sub-population densities, $f_{t,k}$, have partial derivatives up to order k^*, satisfying a Lipschitz condition uniformly on a compact subset, $\mathcal{X} \subset \mathbb{R}$, for*

$t = 1, \ldots, T$ *and* $k = 1, \ldots, L$.

Assumption 2. *1) The kernel weight* $K(u)$ *used in the estimation procedure is a bounded probability density function with support,* $\{u : |u| \leq 1\}$, *such that* $K(u) = K(-u)$, *and* $\int K^2(u)du < \infty$. *2) The weighting function* $w(\cdot)$ *in the inner product* $\langle \cdot, \cdot \rangle$ *is a bounded continuous non-negative function supported on an open convex set of the real line.*

Assumption 3. *There exist constants,* $0 < C_{3,r} < \infty$ *and* $0 < C_{4,r} \leq C_{5,r} < \infty$, *such that* $\min_{s=1,\ldots,T; s \neq r} |\lambda_r - \lambda_s| \geq C_{3,r}T$ *and* $C_{4,r}T \leq \lambda_r \leq C_{5,r}T$ *for any* T, N, *and any fixed* $r \in \{1, \ldots, L\}$.

Assumption 4. *For each* $k \in [1, L]$ *and* $i \in [1, N_k]$, *the random variables* $X_{k,i}(t)$ *are strongly mixing in the sense of Rosenblatt (1956) with the strong-mixing coefficient* $\alpha(\tau)$ *verifying* $\sum_{\tau=1}^{\infty} \tau\alpha(\tau) < \infty$.

Assumption 1 is a stronger version of Assumption A1 in Kneip and Utikal (2001) in that it requires all sub-population densities and their partial derivatives to satisfy a uniform Lipschitz condition. Assumption 2 is standard in the kernel smoothing literature, while Assumption 3 is the same as A2 in Kneip and Utikal (2001). Assumption 4 is introduced here to account for possible serial dependence among elements in the sub-populations in the form of a strongly mixing process. Note that a natural implication of this assumption is that the random probability densities of the random variable $X_{k,i}(t)$ are also serially correlated. However, as pointed out by Kneip and Utikal (2001) and Benko et al. (2009) the proposed method still work, for dependent random densities as long as $\langle f_t - f_\mu, \xi \rangle (f_t - f_\mu)$, where $\xi \in \mathcal{L}^2(D)$ with $D \subset \mathbb{R}$ forms a stationary ergodic process. In our application, a simple visual inspection of the estimated principal components in Figure 5 confirms the latter.

The following theorem effectively adapts Theorem 1 in Kneip and Utikal (2001) to allow for sampling weights and weak dependence across samples $S(1), \ldots, S(T)$:

Theorem 1. *Let* $h_k = O(N_k^{-\delta_k})$ *for some* $\delta_k \in [1/4, 1)$, $k = 1, \ldots, L$, *such that* $(T^{1-4\epsilon}/h_k^2)$ $\alpha\left(T^{1/2+\epsilon}\right) \downarrow 0$ *and* $(\sqrt{N^\wedge T}/h_k)\alpha\left(T^{1/2-\epsilon}\right) \downarrow 0$ *for some* $\epsilon \in (0, 1/2)$. *Suppose that* $T/N^\wedge \downarrow 0$. *Then,*

1. *For a fixed* $r \in \{1, \ldots, T\}$,

$$\widehat{\lambda}_r - \lambda_r = p_r^\top (\widehat{\mathcal{M}} - \mathcal{M}) p_r + O_p(T/N^\wedge) = O_p(\sqrt{T/N^\wedge}). \tag{14}$$

2. *Write* $q_r = (q_{1,r}, \ldots, q_{T,r})$ *with*

$$q_{t,r} = \sum_{s=1}^{T} p_{s,r}(\langle \widehat{f}_{n,s} - f_s, f_t - f_\mu \rangle + \langle \widehat{f}_{n,t} - f_t, f_s \rangle - \langle \widehat{f}_{n,\mu} - f_\mu, f_s \rangle)$$

and $S_r = \sum_{\substack{s=1 \\ s \neq r}}^{T} \frac{1}{\lambda_s - \lambda_r} p_s p_s^\top$. *One has that*

$$E\left[\left|\widehat{p}_r - p_r - S_r(\widehat{\mathcal{M}} - \mathcal{M})p_r\right|\right] = O_{a.s.}\left(\frac{1}{N^\wedge}\right) \tag{15}$$

and

$$E\left[\|\widehat{p}_r - p_r - S_r q_{t,r}\|\right] = O_{a.s.}\left(\frac{1}{N^\wedge} + \frac{1}{N^\wedge\sqrt{Th^\wedge}} + \frac{(h^\vee)^2}{\sqrt{TN^\wedge}}\right). \tag{16}$$

Moreover, we also have $|\widehat{\theta}_{t;r} - \theta_{t;r}| = O_p\left(N^{\wedge -1/2}\right)$ *and* $\|\widehat{p}_r - p_r\|_2 = O_p\left(N^{\wedge -1/2}\right)$.

The asymptotic normality of the normalized partial summation of eigenvalues $\sqrt{N^\wedge/T}\sum_{r=J+1}^{T}\widehat{\lambda}_r$ is provided in Section B in the supplemental material (Chu et al., 2018). Theorem 1 shows that the speed of convergence of the estimated eigenvalues and strength components to their true values is governed by how quickly the number of observations in the smallest strata increases. Our proposed bias correction that adjusts for possible serial dependence in the raw data enables one to obtain the same convergence rates for the biases of $\widehat{\lambda}_r$, $\widehat{p}_r$, and $\widehat{\theta}_{t;r}$, $r \in \{1, \ldots, T\}$, (see, i.e., (14) (16) above), as in Kneip and Utikal (2001, Theorem 1).

Since the right-hand side of (13) is just a weighted sum of T SWKD estimators, the following theorem characterizes the pointwise asymptotic normality of the functional principal components:

Theorem 2. *Define* $\theta^*_{t;r} = \theta_{t;r}/\sum_{t=1}^{T}\theta^2_{t;r}$ *and suppose that bandwidths* $b_k \downarrow 0$ *such that* $n_k^{1/2}b_k^{5/2} \to$ const., $b_k = O(b_\ell)$ *for every* $k \neq \ell$, *and* $n_k b_k \uparrow \infty$. *Moreover, let* b_k *satisfy* $\frac{N^\vee}{b^\wedge T^{4\epsilon}}\alpha\left(T^{1/2+\epsilon}\right) \downarrow 0$ *and* $\sqrt{\frac{N^\vee}{b^\wedge}}\alpha\left(T^{1/2-\epsilon}\right) \downarrow 0$ *for some* $\epsilon \in (0, 1/2)$. *Under the assumptions of Theorem 1, it then holds that:*

$$\sqrt{N^\vee b^\vee}(\widehat{g}_r - g_r) \xrightarrow{d} N\left(\lim_{N\uparrow\infty}\sqrt{N^\vee b^\vee}\frac{1}{2}\left\{\int x^2K(x)dx\right\}\sum_{k=1}^{L}W_k b_k^2\sum_{t=1}^{T}\theta^*_{t;r}f''_{t,k}(x),\right.$$
$$\left.\lim_{N\uparrow\infty}N^\vee b^\vee\sum_{k=1}^{L}W_k^2\frac{1}{n_k^2 b_k}f_{t,k}(x)\sum_{t=1}^{T}\sum_{i=1}^{N_k}I_{k,i}(t)\theta^{*2}_{t;r}\int K^2(x)dx\right). \tag{17}$$

As in traditional density estimation, this theorem highlights that the functional components will display less variability in the tails of the distribution of the data than in its interior, and that larger biases are associated with more curvature. Note that the bandwidth sequences are allowed to differ across clusters, as well as across different estimated eigenfunctions. Furthermore, our simulation study below shows that the asymptotic approximation can be quite accurate even with effective samples of just a few hundreds from stratas with just a couple of thousands observations to draw from. Since the functional principal components estimators are asymptotically normal, one can calculate pointwise confidence intervals using the bootstrap procedure described in Section 5.3.

For our dataset, it can be argued that, at each point in time, prices set by nearby stores tend to be correlated. This spatial dependence in the prices can effectively be modeled by a strongly mixing random field. The strong-mixing condition quantifies the notion of weak dependence in random fields. In the simplest case, one can assume that prices in each stratum and at each point in time are generated by an m-dependent random field while maintaining that prices in different strata are cross-sectionally independent.

In particular, let $\boldsymbol{i}$ represent the 'site' (*or* location) of a store, i, and V_k is the collections of all sites in a stratum, k. A close inspection of the term $\widehat{\tau}_{ts}$ (defined in the supplemental material (Chu et al., 2018)) reveals that the source of biases in the cross-sectional dimension lies in the summands close to the diagonal in the double summations of type $\sum_{\boldsymbol{i}\in V_k}\sum_{\boldsymbol{j}\in V_k}$. Therefore, a bias-corrected estimate of $M_{t,s}$ would involve the truncated double summations of type $\sum_{\boldsymbol{i}\in V_k}\sum_{\boldsymbol{j}\in V_k:\ \|\boldsymbol{j}-\boldsymbol{i}\|>m}$ where $\|\cdot\|$ is the Euclidean norm–that appear in $\widehat{\tau}_{ts}$.

Moreover, Tran (1990) showed that the kernel density $\widehat{f}_{n,t}$ remains asymptotically normal with the same asymptotic variance as in the independence case if the spatial dependence is sufficiently weak and the bandwidth tends to zero at some rate depending on the decay rate of the strong-mixing coefficient. Therefore, we conjecture that the asymptotic properties of the estimates of the principal components and the score functions (calculated from the aforementioned bias-corrected estimate of the variance-covariance matrix $\widehat{\mathcal{M}}$) still remain valid. However, the formal proof of this conjecture would involve many lengthy and non-trivial arguments and therefore is left for future research.

Finally, we can see in view of Theorems 1 and 2 above and Proposition 1 in the supplemental material (Chu et al., 2018), that if the serial dependence is sufficiently weak, the sequence of bandwidths in each cluster can be chosen as originally suggested by Kneip and Utikal (2001). This is illustrated both in our simulation study and in the empirical applications in the next sections.

4 Simulation Study

In this section we conduct simulations to assess the performance of the proposed estimators of $\{\{\theta_{t,j}\}_{j=1}^{J}\}_{t=1}^{T}$ and $\{g_j\}_{j=1}^{J}$ with stratified random samples using various scenarios, sample and population sizes, and weighting schemes.

We consider the case when $J = 1$ and $T = 30$, where the super-populations $\{f_t\}_{t=1}^{30}$ are generated by a mixture of two normal distributions, i.e., $f_t = f_\mu + \theta_{t,1}g_1$, where $g_1 = \phi_{-1/2,1/2} - \phi_{1/2,1/2}$, with $\phi_{\mu,\sigma}$ representing a Normal density function with mean μ and variance σ^2. The sequence $\theta_{t,1} = \vartheta_t - \vartheta$ and function $f_\mu = \vartheta\phi_{-1/2,1/2} + (1-\vartheta)\phi_{1/2,1/2}$, where $\vartheta = (1/30)\sum_{t=1}^{30}\vartheta_t$, are generated using three scenarios as follows:

Scenario 1: $\vartheta_t = \frac{t}{30} + \frac{1}{4\pi}\sin\left(\frac{\pi t}{30}\right) - \frac{3}{8\pi}\sin\left(\frac{2\pi t}{30}\right)$,
Scenario 2: $\vartheta_t = \frac{1}{2} + \frac{1}{2}\sin\left(\frac{30}{t}\right)$,
Scenario 3: $\vartheta_t = \frac{1}{2}\text{Beta}_{3/2,3/2}(\frac{t}{30})$,

where $\text{Beta}_{\alpha,\beta}$ represents the Beta probability density function with shape parameters α and β. The implied dynamics of these super-population densities are shown in Figure 2. Note that these three scenarios provide rich dynamic settings in which to assess the numerical performance of the proposed estimators in this paper. Scenario 1 provides a setting in which densities are diverging systematically from each other, while Scenario 2 depicts a situation where densities seems to be shuffling around at first and then a pattern of convergence emerges. Finally, Scenario 3 portrays a case in which densities differ very little before and after a period of large changes in the dynamics of these

super-populations.

At each period, t, we then randomly draw (without replacement) a total sample of $n = \sum_{k=1}^{6} n_k$ from $N = \sum_{k=1}^{6} N_k$ observations distributed over $L = 6$ strata as described in Table 1.

Table 1: Simulation Design

n_1/N_1	n_2/N_2	n_3/N_3	n_4/N_4	n_5/N_5	n_6/N_6	n/N
667/2,223	223/4,445	1067/6,666	555/2,219	334/2,223	201/2,224	3,047/20,000
1,501/3,334	501/6,667	2,400/9,999	1,250/3,331	751/3,334	451/3,335	6,854/30,000
2,667/4,445	889/8,889	4,267/13,333	2,221/4,442	1,334/4,445	801/4,446	12,179/40,000

To simplify calculations and computational time we set $f_{t,1} = \ldots = f_{t,6} = f_t$ in (7). For the same reasons bandwidths are also set as $h_1 = \ldots = h_6 = h$, $b_1 = \ldots = b_6 = b$, $h = \widehat{h}^{5/4}$, and $b = \widehat{h} \times T^{-1/5}$, where $\widehat{h}$ corresponds to Duin's (1976) likelihood cross-validation bandwidth. These are done in each period $t = 1, \ldots, 30$; and along with a Gaussian kernel, the SWKD estimators are also estimated, i.e., $\widehat{f}_{n,1}, \ldots, \widehat{f}_{n,30}$ and $\widehat{f}^{*}_{n,1}, \ldots, \widehat{f}^{*}_{n,30}$. We then proceed to calculate $\{\widehat{\theta}_{t,1}\}_{t=1}^{30}$ and $\widehat{g}_1$ on an equally spaced grid on $[-2.7, 2.7]$ using three different weighting functions (see Figure E.1 in the supplemental material (Chu et al., 2018)):

$$w_1(u) = 1, \quad w_2(u) = \frac{10}{\pi[\exp(u) + \exp(-u)]}, \quad w_3(u) = \frac{16|u|}{\pi[\exp(u) + \exp(-u)]}. \tag{18}$$

In (18), w_1 provides equal weights to all points in the integral, while w_2 downweights extreme points in favor of points near zero, and w_3 downweights most points in the center of the distribution, while giving higher weights to points in the tails.

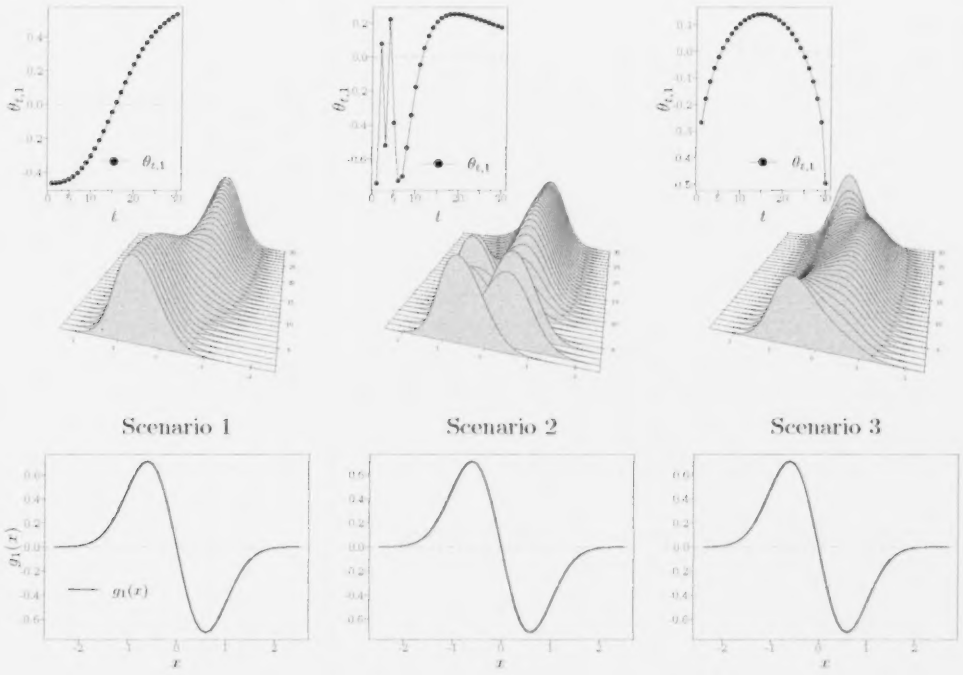


Figure 2: Super-populations Used in the Simulation Study. Top row panels display sequences $\{\theta_{t,1}\}_{t=1}^{30}$ that, along with the common g_1 function displayed in the bottom row panels, generate the set of 30 super-populations, $f_1, \ldots, f_{30}$, displayed in the middle row panels. The dashed horizontal lines mark zero.

The experiment is repeated 1,000 times and Figure 3 summarizes our results (the supplemental material (Chu et al., 2018) shows the results when using Silverman's (1986) rule-of-thumb bandwidths instead). Figure 3 displays the simulated Root Mean Squared Error (RMSE) of estimators of $\theta_{10,1}, \ldots, \theta_{20,1}$ in each scenario when using weighting functions w_1, w_2, and w_3 in (18). These simulations show that the precision of the proposed estimators for the dynamic strengths components increases with sample and population sizes when accounting for the complex survey design. The performance of the estimators is, however, sensible to the weighting functions being used. In particular, the performance in terms of RMSE using weighting schemes w_1 or w_2 is strictly dominated by the weighting function w_3 in Scenarios 1 and 2. On the other hand, although weighting scheme w_3 displays smaller RMSE than when using w_2, no weighting scheme w_1 gives the smallest simulated RMSE in Scenario 3. Interestingly, as seen in Figure 2, weighting function w_3 seems to work best when the intertemporal dynamics concentrates in the tails, and these are given larger weights as can be seen in Scenarios 1 and 2.

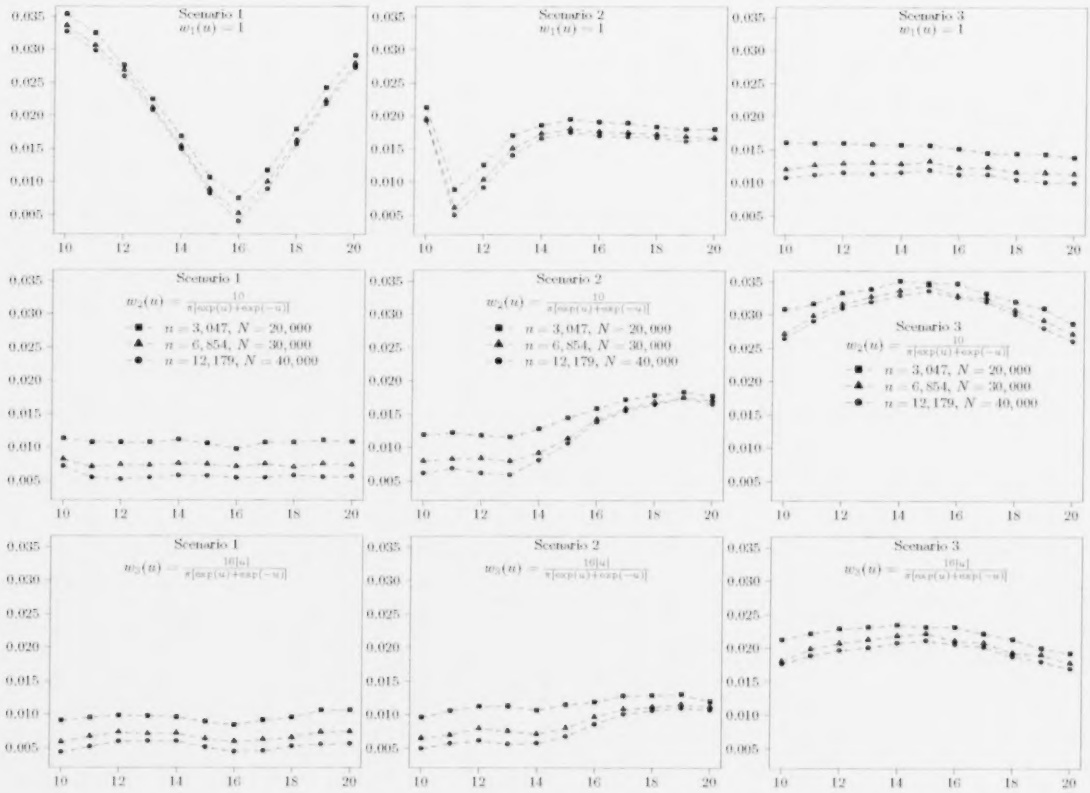


Figure 3: Simulated Root Mean Squared Error of the Dynamic Strength Components. Simulated Root Mean Squared Error of estimators of $\theta_{10,1}, \ldots, \theta_{20,1}$ in various sampling designs for all three scenarios using all three weighting functions.

Similarly, Figure 2 in the supplemental material for this paper (Chu et al., 2018) shows the Q-Q plots of the standardized simulated $\widehat{g}_1$ evaluated at $x = 0.5$. These plots show that the pointwise asymptotic normal approximation is good across all scenarios and sample/population sizes. Again, the performance of the weighting function in w_3 seems to dominate throughout in terms of tail behavior. These results along with those in Figure 3 are in line with our theoretical predictions in Section 3.2. They also highlight the fact that the weighting function w_3 seems to provide better estimates of the strengths components and basis functions. Furthermore, as shown in the supplemental material (Chu et al., 2018), these findings are robust to bandwidth choice.

5 Applications to the UK CPI Data

In this section, we demonstrate that the FPCA methodology can be used to visualize in a structured manner the evolution of micro-price distributions, see, e.g., Huynh and Jacho-

Chávez (2010) and Huynh et al. (2016) for similar applications. We apply the FPCA methodology to the UK price data to analyze the time evolution of the distribution of the standardized and detrended natural-logarithm prices (log(Price)) and their changes (Δ log(Price)). We utilize the `R` packages `npRmpi` (cross-validation) and `snow` (numerical integration). All computations were performed on EDITH, the Bank of Canada High Performance Cluster, which consists of 664 cores on 36 nodes. All jobs utilize about 250 cores; bandwidth cross-validation takes about 24 hours while numerical integration takes about an hour.

5.1 Data Visualization

Measurement of the dynamics of economy-wide price distributions offers important implications for economists and central bankers, who are interested in understanding fluctuations in inflation and other economic variables. Understanding the evolution of the joint dynamics of micro-level prices is of first-order importance for economists. A voluminous literature in this area is represented by papers including inter alia Bils and Klenow (2004), Dhyne et al. (2006), Klenow and Kryvtsov (2008), Nakamura and Steinsson (2008), Gagnon (2009), Klenow and Malin (2010), Kryvtsov (2016), and Kryvtsov and Vincent (2017). Since inflation measures a change in the price level for goods and services sold by retailers across the country, inflation dynamics depend on how individual retailers set and change their prices across time, and on how price behavior across retailers balances out the economy-wide scale. Information about these two sources of aggregate price movements is handily embedded in price distributions and their movements across time. Micro-price data can be used to estimate price distributions, and therefore, provide valuable evidence on the nature of aggregate inflation fluctuations and the effects of monetary policy.

Figure 4 utilizes Hyndman et al.'s (1996) highest density regions, stacked density plots, and the smoothed version in Castro Camilo and de Carvalho (2017) to display the resulting SWKD estimators for both variables using a second-order Gaussian kernel in each month. Given the large number of observations and the absence of sparseness within strata, we use a common bandwidth across stratas that are chosen by Duin's (1976) likelihood cross-validation. This cross-validation is completed for each month in our sample and the resulting bandwidths can be found in the top graph in Figure E.4 in the supplemental material (Chu et al., 2018). We observe several empirical regularities for the distributions of the demeaned price levels (log(Price), left panel) and price changes (Δ log(Price), right panel).

First, from month to month, distributions shift their mass across different density regions, captured by shaded areas on the right-hand-side plots. Second, the relative sizes of density regions change across time, reflecting fluctuations in the degree of price dispersion. Third, changes in relative density regions are asymmetric from month to month, resulting in varying skewness of the distribution (top- and middle-left panels). Fourth, the densities at the peak and at the tails vary extensively (middle- and bottom-right panels). Finally, fluctuations in density regions stand out during recessions. For

example, the peak of price-level distributions is markedly higher during mid-2000s, and the 50% density is smaller for price changes during both recessions.

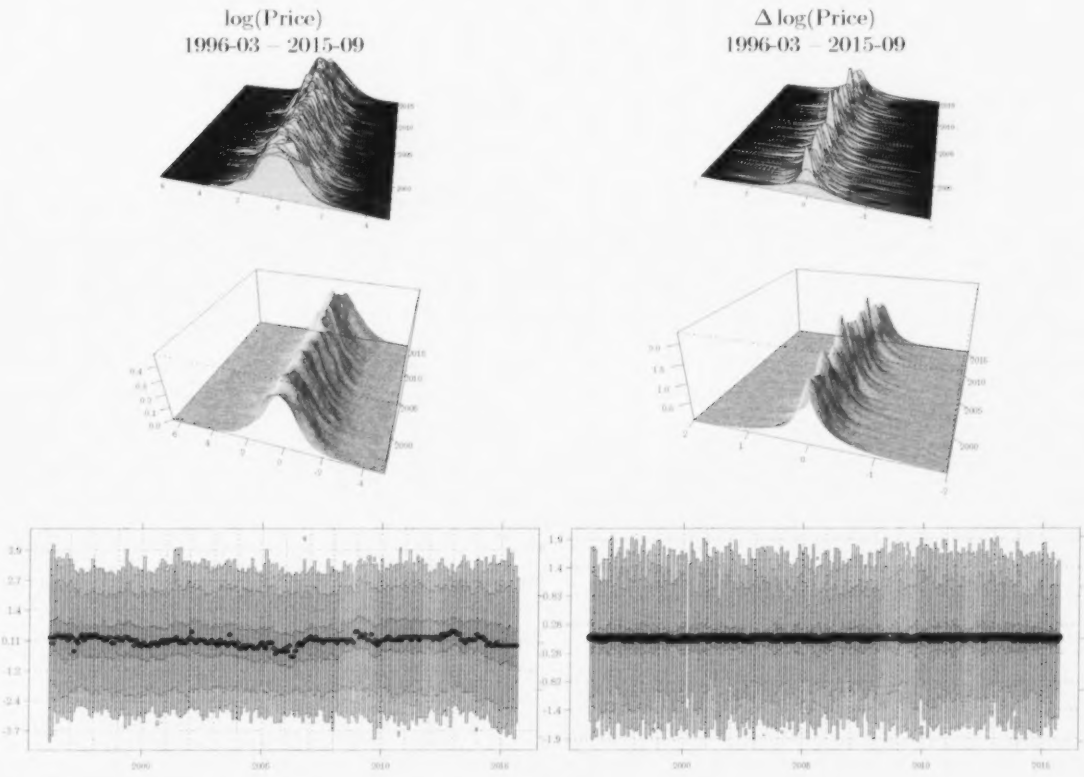


Figure 4: Estimated Distributions of log(Price) and Δ log(Price). Top and bottom graphs utilize Hyndman et al.'s (1996) stacked density plots device and their highest density regions per month with 50% (dark gray), 95% (gray), 99% (light gray) probability coverage, and the mode (•). Middle plots display Castro Camilo and de Carvalho's (2017) smooth stacked density plots.

These descriptive statistics are visually appealing. However, it is difficult to quantify how these distributions are evolving over time, especially with complex survey weights. In the next section, we operationalize the estimators described in Section 3.1.1 by using weighting function (w_1) in (18) over the observed support of the entire sample (as permitted in Assumption 2). This weighting function was found to outperform *no weighting* in the simulation study in Section 4. The required bandwidths in each month are chosen as described in Section 3.1.1. A total of $T \times (T-1)/2 = 27,495$ univariate numerical integrals were calculated in parallel using an adaptive quadrature algorithm in Piessens et al. (1983). Results are shown in Figure 5.

5.2 Dynamic Scree Plots and Strength Coefficients

The left-hand-side plots in Figure 5 display the dynamic scree plots–the contribution of the respective eigenvalues to total variation of distributions. Visual inspection of the dynamic scree plot for the log(Price) reveals that the first four components account for about 5% of the log(Price) distributions variation after which the dynamic scree plot is relatively flat. For the Δlog(Price) the first four components account for 7% of the variation. Hence, a bulk of the time-variation in price distributions is characterized by just a few principal components. Furthermore, the scree plots demonstrate that the second, third, and fourth components together are about as important as the first common component in accounting for the variance of these distributions. This is an important result for understanding the drivers of inflation fluctuations as it highlights the key role of higher order components.

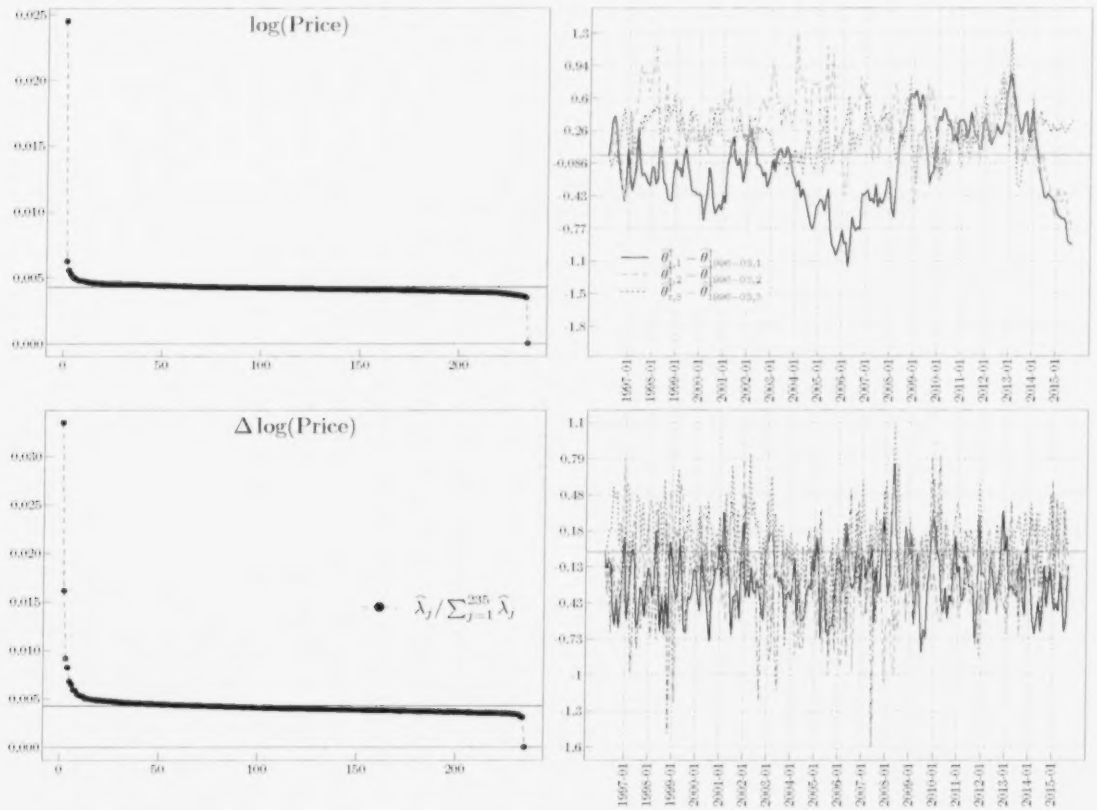


Figure 5: Estimated Dynamic Scree Plots and Strength Components: log(Price) and Δ log(Price). Each of the plotted estimated strength components is normalized to start at zero on 1996–03, i.e., $\widehat{\theta}^{\dagger}_{t,j} - \widehat{\theta}^{\dagger}_{1996-03,j}$, where $\widehat{\theta}^{\dagger}_{t,j} = \widehat{\theta}_{t,j} / \min_t \widehat{\theta}_{t,j}$, for $j = 1, 2, 3$.

The right-hand side of Figure 5 provides a normalized version of the estimated first

three dynamic strength coefficients, i.e., $\widehat{\theta}^{\dagger}_{t,r} - \widehat{\theta}^{\dagger}_{1996-03,r}$, where $\widehat{\theta}^{\dagger}_{t,r} = \widehat{\theta}_{t,r} / \min_t \widehat{\theta}_{t,r}$, and $\widehat{\theta}_{t,r}$ is defined in (13), for $r = 1, \ldots, 3$. These plots describe how each component evolves over time. Note that these location/scale normalizations are useful in describing the movements in terms of a common initial value of zero, and in terms of order-of-magnitudes larger than the lowest temporal deviation, i.e., scale free. They also better highlight the correlation between them and other key economic variables. The dynamics of the deviations in the first three estimated dynamic strength coefficients for the log(Price) resemble an AR(1) process with the AR estimated coefficients of 0.19, 0.61, and 0.64, respectively. For the Δlog(Price) the process has AR(1) coefficients of 0.37, 0.22, and -0.36, respectively.

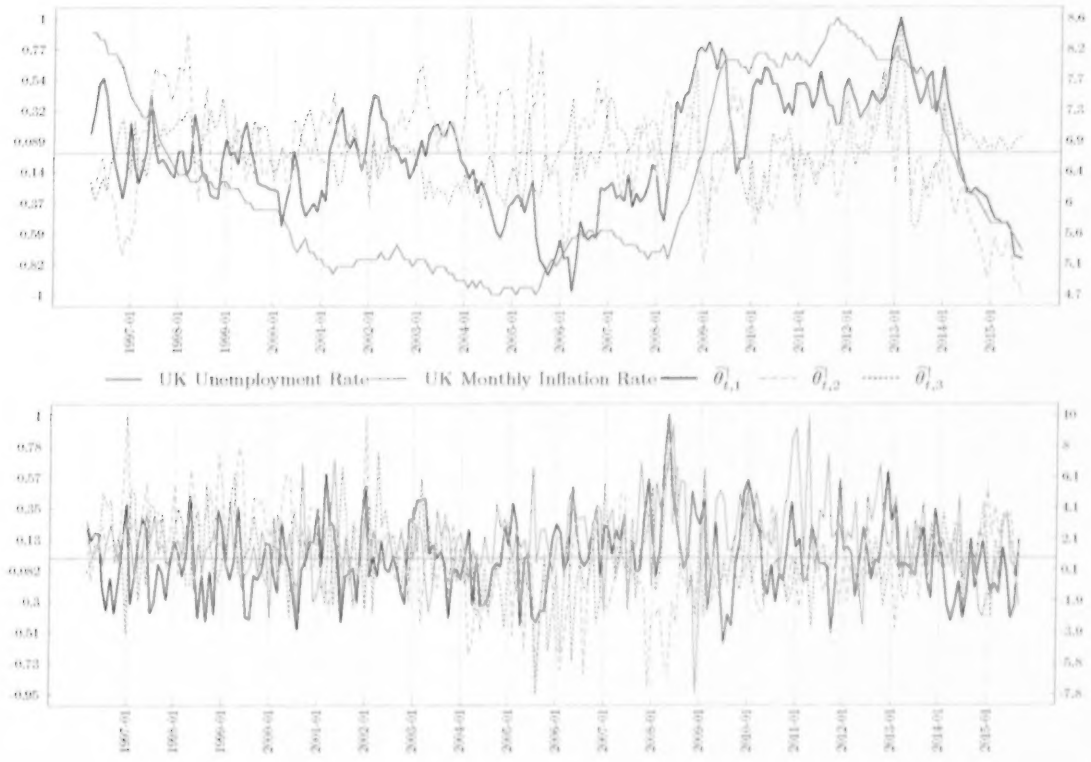


Figure 6: Estimated Strength Components, UK Unemployment, and Inflation. Top plot: The left-hand-side axis is the scale for the normalized estimated strength components, i.e., $\widehat{\theta}^{\dagger}_{t,j} = \widehat{\theta}_{t,j} / \min_t \widehat{\theta}_{t,j}$, $j = 1, 2, 3$, for log(Price) (red solid line represents its zero), while the right-hand-side axis is the scale for the UK monthly unemployment rate measured in percentages. Bottom plot: The left-hand-side axis is the scale for the normalized estimated strength components (defined before) for $\Delta \log(\text{Price})$, while the right-hand-side axis is the scale for the UK monthly inflation as measured by the CPI measured in percentages.

We find that some of these components are correlated with economic variables. Figure 6 illustrates that the first three estimated strength coefficients for log(Price) and Δlog(Price) comove with the UK monthly unemployment rate and the inflation rate. For log(Price), the correlations of the first three estimated dynamic strength coefficients and the unemployment rate are 0.66, -0.20, and 0.12, respectively; and for Δlog(Price), the correlations of the first three estimated dynamic strength coefficients and the inflation rate are 0.20, -0.11, and -0.06, respectively.

While we are able to estimate the correlations consistently, their standard errors remain unknown so we are unable to test their statistical significance. Nonetheless, since

by construction the dynamic strength coefficients are statistical 'model-free' concepts, correlation of the coefficients with economic variables suggests that FPCA methodology has potential value for price measurement and macroeconomic analysis. Note that these correlations already take into account the information contained in the sampling design through the estimated coefficients as described in Theorem 2 above.

5.3 Basis Functions and their Deformations

The estimated dynamic strength coefficients summarize the dynamics of the distributions through time, but they do not provide information about the parts of the distributions that are evolving. This information is captured by the estimated time-invariant basis functions, or functional principal components, $\widehat{g}_j$. Black solid lines in Figures 7 and 8 plot the first four bootstrap bias-corrected estimated common functional components, $\widehat{g}_j$, $j = 1, \ldots, 4$, for log(Price) and Δlog(Price), respectively.

The shape of the common functional components outlines the density regions that are deformed over time, and the extent of the temporal deformation is given by the corresponding estimated dynamic strength coefficients, $\widehat{\theta}_{t,j}$. For example, $\widehat{g}_1$ on Figures 7 and 8 is positive for prices and price changes that are roughly below zero and negative for prices and price changes that are above zero, respectively. This means that time variation in the estimated dynamic strength coefficient $\widehat{\theta}_{t,1}$ will be shifting the mass in the distribution of prices (price changes) back and forth from low to high prices. The mass that is distributed is given by the total estimated component deformation, $\widehat{\theta}_{t,1}\widehat{g}_1$. Likewise, estimated higher order basis functions represent mass shifts away from the middle ($\widehat{g}_2$), between the tails ($\widehat{g}_3$), and between the middle and the tails ($\widehat{g}_4$). These inherent features of the first four common functional components can therefore be related to the temporal variations in the first four central moments of the distributions: the mean, variance, skewness, and kurtosis.

To illustrate the usefulness of FPCA, Figures 7 and 8 plot the estimated common component deformations, $\widehat{\theta}_{t,j}\widehat{g}_j$, for three important periods of the British economy in the past 19 years: for April 2001, representing an expansionary period of economic activity ($\widehat{\theta}_{2001-04,j}\widehat{g}_j$ for $j = 1, 2, 3, 4$), and the beginning of two recessions in August 2009 ($\widehat{\theta}_{2009-08,j}\widehat{g}_j$) and December 2011 ($\widehat{\theta}_{2011-12,j}\widehat{g}_j$). The unemployment rate was 5% in April 2001 and 8% (8.5%) in August 2009 (December 2011).

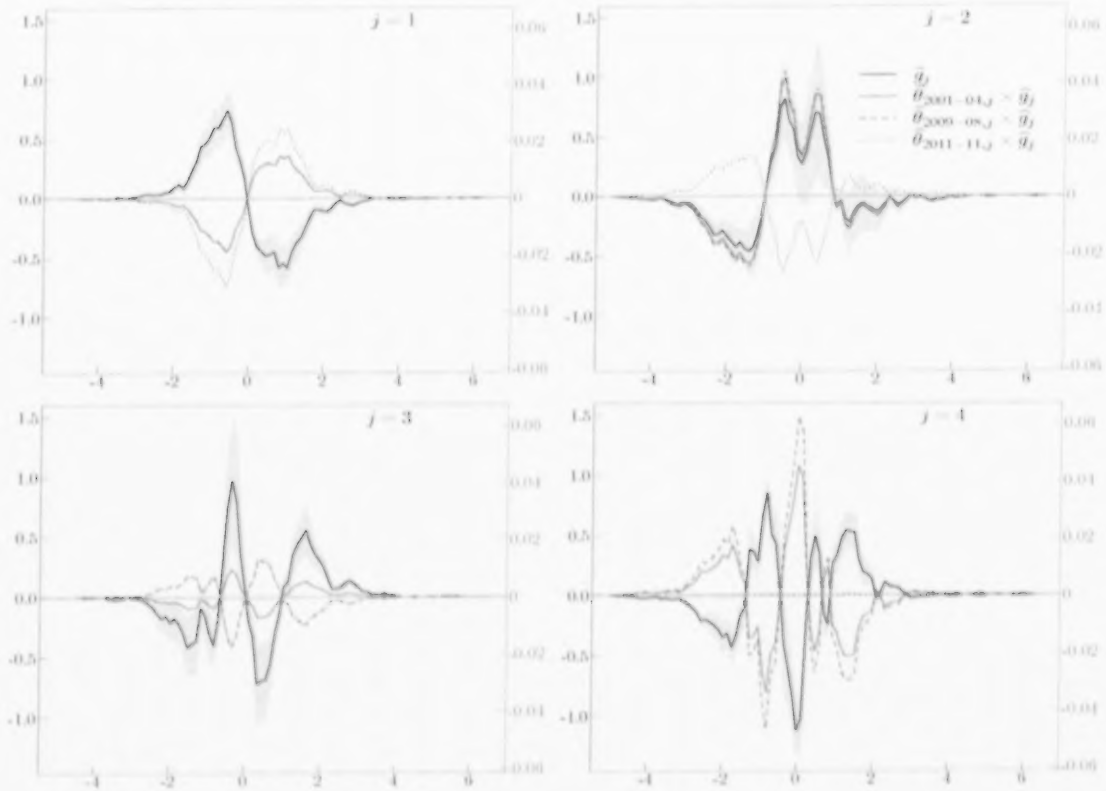


Figure 7: Estimated First Four Basis Functions and their Deformations for log(Price). Each plot displays bootstrap bias-corrected estimates of the first four basis functions, $\widehat{q}_j(\cdot)$, for $j = 1, \ldots, 4$ (black solid line), as well as three deformations corresponding to the beginnings of two recessions, i.e., 2009–08 (blue dashed line) and 2011–11 (blue dotted line), as well as the beginning of a period of normal economic activity, i.e., 2001–04 (blue solid line). The red solid line represents zero.

Based on the point-wise asymptotic normality result in Theorem 2, 95% bootstrap point-wise confidence intervals based on 999 replications are also shown in these plots, as gray areas. To mimic the original stratified sampling design in each month, the following bootstrap procedure was implemented:

1. For each month, $t = 1, \ldots, 235$, 999 bootstrap samples within strata are taken and bootstrap replication weights are constructed as suggested in Canty and Davison (1999, Section 3.4, pp. 383-384).

2. Based on the 999 bootstrap samples and replication weights generated in the first step, bootstrap analogues of the original estimator of (13) are then constructed using bandwidths and kernel functions used for the original sample.

Hence, components with basis functions for which gray areas are above or below zero make a statistically significant impact on the underlying distributions.

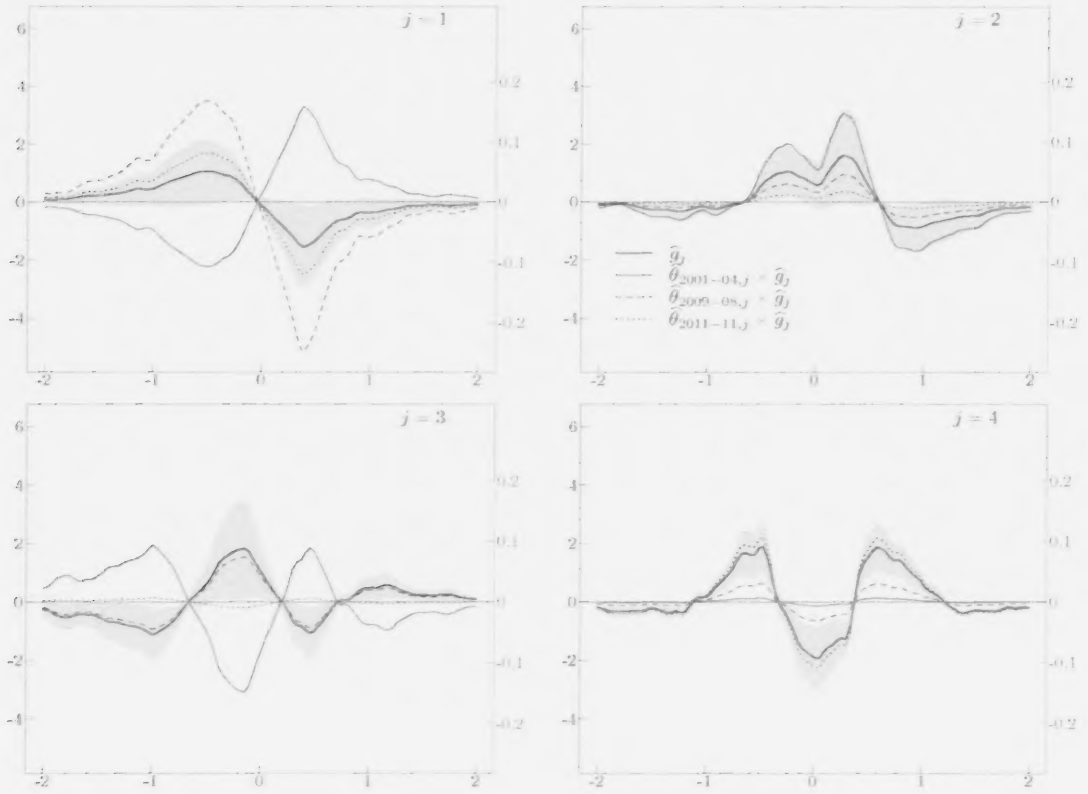


Figure 8: Estimated First Four Basis Functions and their Deformations for $\Delta \log(\text{Price})$. Each plot displays bootstrap bias-corrected estimates of the first four basis functions, $\widehat{g}_j(\cdot)$, for $j = 1, \ldots, 4$ (black solid line), as well as three deformations corresponding to the beginnings of two recessions, i.e., 2009–08 (blue dashed line) and 2011–11 (blue dotted line), as well as the beginning of a period of normal economic activity, i.e., 2001–04 (blue solid line). The red solid line represents zero.

The first estimated common component deformation shows that in 2011 the underlying baseline distribution of prices shifts the weight from the left part of the distribution (less than zero) toward the right part of the distribution, which is indicative of an increase in the mean. Note that the 2001 and 2009 estimated common component deformations are not discernible in the figure. The second component deformation for 2009 (top-right panel in Figure 7) indicates that the weight is moving toward the center of the distribution, which is indicative of decrease in price distribution variance in the wake of the Great Recession in the UK. The shape of the corresponding estimated basis function

(bottom-left panel in Figure 7) implies that during slumps there is less weight assigned to the left of the distribution so that positive skewness is lower, which is the case for the 2009 and 2011 recessions and is virtually not discernible on the figure. The fourth estimated basis function (bottom-right panel) complements this interpretation by showing that the middle part of the distribution is gaining at the expense of the tails.

Similarly, for $\Delta\log(\text{Price})$ in Figure 8, the first, second, and third components' confidence intervals contain zero for some of the range, but are nonetheless informative. For example, the first component reveals a positive mean-effect in 2001 while in 2009 there is a negative mean-effect, which is consistent with predictions of standard sticky price models that economic expansions (contractions) are associated with higher (lower) inflation rates. The second estimated common component deformations (top-right panel in Figure 8) suggest that the 2009 and 2011 recessions are associated with larger dispersion than the 2001 expansion. The third estimated common component deformations show an increase in the left tail at the expense of the middle and right tail in the 2001 expansion, which is indicative of higher positive skewness. The fourth estimated common component deformation shows that during the 2009 and 2011 recessions, the mass in both tails and the middle of the distribution was lower, which is indicative of higher kurtosis.

5.4 Density Forecasting

In this section, we propose a density forecast exercise based on (1) to forecast price distributions and track the underlying price dynamics. This exercise also serves as a data-driven mechanism for choosing J, the number of principal components. First, recall we can estimate nonparametrically T super-populations, namely $f_1, \ldots, f_T$, based on a sample $\{\{X_{it}, w_{it}\}_{i=1}^{n_t}\}_{t=1}^{T}$ as in (8), i.e., $\{\widehat{f}_t\}_{t=1}^{T}$. By choosing $T^* < T$, one can implement the following procedure to select the number of components, J, and the best (in the Integrated Squared Error sense) forecast, $\hat{f}_{T^*+\ell|T^*}$ for $\ell = 1, 2, \ldots$:

Step 1. Using the first T^* estimated densities, $\{\widehat{f}_t\}_{t=1}^{T^*}$, calculate $\{\{\widehat{\theta}^*_{t,r}\}_{r=1}^{L^*}\}_{t=1}^{T^*}$ and $\{\widehat{g}^*_r\}_{r=1}^{L^*}$, where $L^* \leq T^*$ represents the number of the first non-zero eigenvalues of the $T^* \times T^*$-matrix $\widetilde{\mathcal{M}}^*$ as in Section 3.1.1. Also set $\widehat{f}^*_\mu = (1/T^*)\sum_{t=1}^{T^*}\widehat{f}_t$.

Step 2. In view of the orthogonality conditions in (5), we utilize the algorithm in Hyndman and Khandakar (2008, Section 3.2, pp. 10–11) to automatically identify the best-fitted ARMA model for each generated series, $\{\widehat{\theta}^*_{t,r}\}_{t=1}^{T^*}$, $r = 1, \ldots, L^*$, and then proceed to obtain an automatic forecast for period $T^* + \ell$ as described in Hyndman and Khandakar (2008, Section 2.6, pp. 8), i.e., $\{\hat{\theta}_{T^*+\ell|T^*,r}\}_{r=1}^{L^*}$.

Step 3. Set

$$\widehat{J} = \arg\min_{l \in \{1,\ldots,L^*\}} \int \left(\widehat{f}_{T^*+\ell}(x) - \widehat{f}^*_\mu(x) - \sum_{r=1}^{l} \hat{\theta}_{T^*+\ell|T^*,r}\widehat{g}^*_r(x) \right)^2 dx, \qquad (19)$$

$$\hat{f}_{T^*+\ell|T^*} = \widehat{f}^*_{\mu} + \sum_{r=1}^{\widehat{J}} \hat{\theta}_{T^*+\ell|T^*,r} \widehat{g}^*_r. \tag{20}$$

Results are displayed in Figure 9. We set $T^* = 232$ (or 2015–06) and calculate $\widehat{J} = 231$ and $\widehat{J} = 4$ for log(Price) and Δlog(Price) respectively. Using these calculated $\widehat{J}$s we then proceed to repeat Steps 1 and 2 above in each of the 999 bootstrap replications described in Section 5.3. The latter is done using the original calculated bandwidths, Gaussian kernel functions, and weighting function (w_1) in (18) in each replication. We then construct 95% point-wise confidence intervals and bias-corrected versions of $\hat{f}_{233}$, $\hat{f}_{235}$, $\hat{f}_{233|232}$, and $\hat{f}_{235|232}$ using these bootstrap samples.

As shown in Figure 9, the one-month ($\ell = 1$) and three-month ($\ell = 3$) population density forecasts for log(Price) and Δlog(Price) are spot-on in terms of overall shape, although the point-wise variability is larger for the one-month forecast density for Δlog(Price). The pointwise bootstrapped confidence intervals are in general tighter for the forecasts than for the actual bias-corrected SWKD estimators. Figure 3 in the supplementary material provides the resulting Q-Q plots for these forecasts as well. The results are discussed there. Alternatively, one can implement various formal quantitative methods for density forecast evaluation as suggested in Thorarinsdottir et al. (2013).

An interesting feature of this density forecast algorithm is that as time passes, it can be re-run to update the chosen $\widehat{J}$ and the estimated parameters of the automated ARMA model for each generated series $\{\widehat{\theta}^*_{t,r}\}_{t=1}^{T^*}$, $r = 1, \ldots, L^*$. However, the theoretical justification of this proposed density forecast algorithm is beyond the scope of the paper and therefore left for future research.

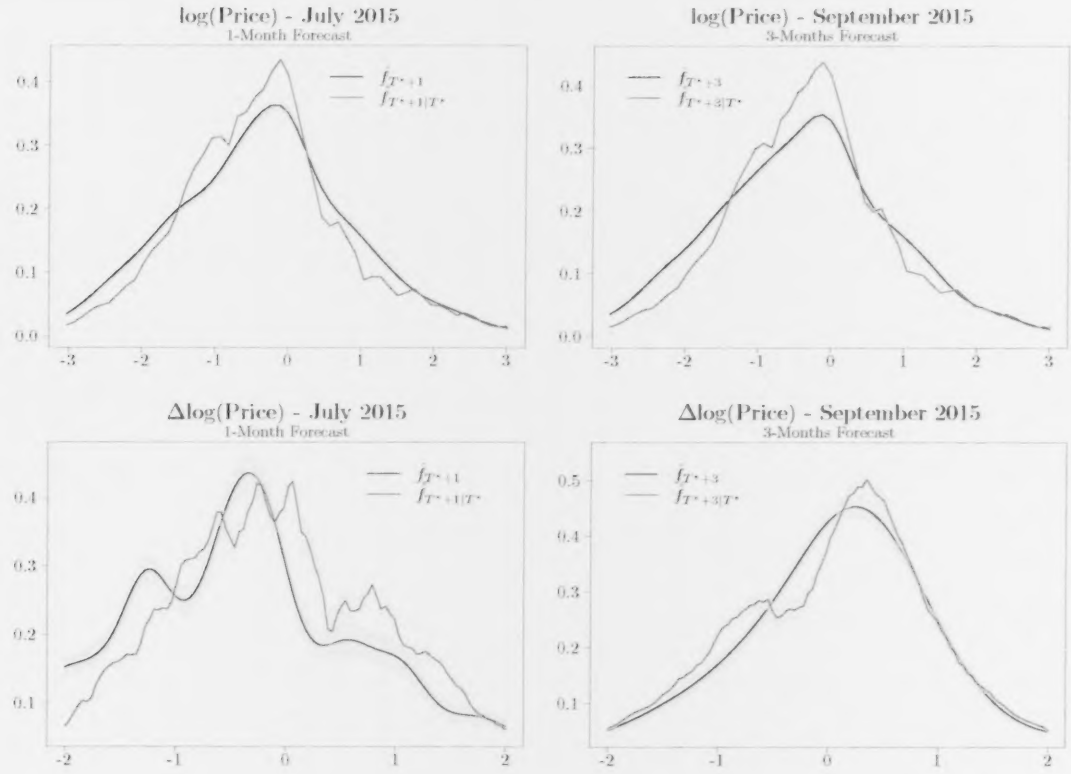


Figure 9: Feasible One-Month and Three-Months Density Forecasts. The light gray and light blue areas represent 95% point-wise bootstrap confidence intervals based on 999 replications.

5.5 Sensitivity Analysis

As a sensitivity analysis, the entire empirical analysis in this paper was performed using naive Silverman's (1986) rule-of-thumb bandwidths instead of Duin's (1976) cross-validated bandwidths; see Section F.5 in the supplemental material (Chu et al., 2018). Similarly, the sensitivity of the results were also analyzed when using the remaining weighting functions in (18) instead; see Section F.4 in the supplemental material (Chu et al., 2018). A simple visual inspection of these results confirms that the empirical findings in this paper are qualitatively the same when using different bandwidths and weighting functions. The general shapes and dynamics remain the same and tell the same story. The density forecasting exercise also seems to be robust to the choice of bandwidths and weighting functions, but the resulting values of $\widehat{J}$ in (19) tend to be lower for Δlog(Price) when using rule-of-thumb bandwidths or the remaining weighting functions in (18) with cross-validated bandwidths instead.

Finally, we also check the sensitivity of the proposed algorithm in Section 5.1 when $\widehat{J}$ in (19) is chosen as $\widehat{J} = \arg\min_{l \in \{1,\ldots,L^*\}} |\widehat{\lambda}_l / \sum_{j=1}^{L^*} \widehat{\lambda}_j - 1/T^*|$, where we set $T^* = 232$, and $L^* \leq T^*$. Results are displayed in Section F.6 in the supplemental material (Chu et al., 2018). We find that this alternative way to choose J provides more parsimonious models when rule-of-thumb bandwidths are used for log(Price). Otherwise, no other clear ranking can be found. Nevertheless, we note that general shapes remain the same indicating that, at least for this application, the choice of J does not seem to play an important role in density forecasting.

6 Discussion

In this paper, we adapt the conventional FPCA method of Kneip and Utikal (2001) to allow for survey weights and temporal dependence. This methodology is useful since most micro data come with survey weights, and it allows an understanding of the joint evolution of these distributions. This method is superior to simple inspection of descriptive statistics tools such as median and quantiles of distribution because we jointly model the cross-sectional and temporal dependence of survey data. In addition, the paper provides asymptotic evidence that our estimator can be adapted to allow for survey weights while correcting for temporal-dependency induced biases. We conduct an extensive simulation exercise with various scenarios that contain different sample and population sizes, and with varying tuning parameters such as weighting functions. The simulation demonstrates the efficacy of our methodology and robustness of our results to the choice of bandwidths and weighting functions. Further, it allows us to decompose how the distribution changes with the respective components.

We highlight the effectiveness of our method by applying it to a large-scale dataset that consists of 26 million unique price quotes used for constructing the UK consumer price index for the period from March 1996 to September 2015. We focus our analysis on the distribution of detrended logarithm of prices, log(Price), and the month-to-month change in the logarithm of prices, Δlog(Price). We find that in our application the dynamics of higher order moments of price and price-change distributions are as important as the first moment in driving these changes over time. For example, price dispersion decreases and the tails thin out in the wake of the 2009 recession, but these effects are reversed over the ensuing 2011 recession. At the same time price changes become more dispersed, suggesting that the times of economic turmoil trigger across-the-board price adjustments. These facts gleaned from this visualization method can be used by economists to sort out models of inflation and business cycles because their predictions are inherently linked to assumptions about individual price adjustments in response to economic shocks.

Finally, we use the FPCA components to conduct an out-of-sample forecasting exercise. We exploit the persistence of the dynamic strength components to compute one-month and three-month out-of-sample forecasts of log(Price) and Δlog(Price). We compare our forecasts with the random walk model and find that log(Price) is observationally

equivalent while the $\Delta\log(\text{Price})$ is superior. This exercise illustrates the usefulness of FPCA for visualization, to decompose the evolution of a distribution, and then to exploit this evolution for forecasting purposes.

7 Supplemental Material

In the supplemental appendix, we provide detailed mathematical proofs of the main theoretical results of the paper. It also contains specific details regarding the numerical methods used in the application such as bandwidths calculation and numerical integration. Moreover, it includes all figures presented in the simulation study and application here, but redrawn using different bandwidths and weighting schemes as discussed in Section 5.5 above.

References

Alvarez, F., Le Bihan, H., and Lippi, F. (2016), "The Real Effects of Monetary Shocks in Sticky Price Models: A Sufficient Statistic Approach," *American Economic Review*, 106, 2817–2851.

Ball, L. and Mankiw, N. G. (1995), "Relative-Price Changes as Aggregate Supply Shocks," *The Quarterly Journal of Economics*, 110, 161.

Bellhouse, D. R. and Stafford, J. E. (1999), "Density Estimation from Complex Surveys," *Statistica Sinica*, 9, 407–424.

Benko, M., Härdle, W., and Kneip, A. (2009), "Common Functional Principal Components," *Annals of Statistics*, 37, 1–34.

Berger, D. and Vavra, J. (2018), "Dynamics of the U.S. Price Distribution," *European Economic Review*, 103, 60–82.

Bils, M. and Klenow, P. J. (2004), "Some Evidence on the Importance of Price Stickiness," *Journal of Political Economy*, 112, 947–987.

Buskirk, T. D. (1998), "Nonparametric Density Estimation Using Complex Survey Data," in *Proceedings of the Survey Research Methods Section*, American Statistical Association, Washington, DC, pp. 799–801.

— (1999), "Using Nonparametric Methods for Density Estimation with Complex Survey Data," Ph.D. dissertation in statistics, Arizona State University.

Buskirk, T. D. and Lohr, S. L. (2005), "Asymptotic Properties of Kernel Density Estimation with Complex Survey Data," *Journal of Statistical Planning and Inference*, 128, 165–190.

Canty, A. J. and Davison, A. C. (1999), "Resampling-Based Variance Estimation for Labour Force Surveys," *The Statistician*, 48, 379–391.

Castro Camilo, D. and de Carvalho, M. (2017), "Spectral Density Regression for Bivariate Extremes," *Stochastic Environmental Research and Risk Assessment*, 31, 1603–1613.

Chen, H. A., Levy, D., Ray, S., and Bergen, M. (2008), "Asymmetric Price Adjustment in the Small," *Journal of Monetary Economics*, 55, 728–737.

Chu, B. M., Huynh, K. P., Jacho-Chávez, D. T., and Kryvtsov, O. (2018), "Supplement to "On the Evoluation of the UK Price Distributions".

Clews, G., Sanderson, R., and Ralph, J. (2014), "Calculating a Retrospective Superlative Consumer Prices Index for the UK," Tech. rep., Office for National Statistics.

Debelle, G. and Lamont, O. (1997), "Relative Price Variability and Inflation: Evidence from U.S. Cities," *Journal of Political Economy*, 105, 132–152.

Dhyne, E., Alvarez, L., Le Bihan, H., Veronese, G., Dias, D., Hoffmann, J., Jonker, N., Lunnemann, P., Rumler, F., and Vilmunen, J. (2006), "Price Changes in the Euro Area and the United States: Some Facts from Individual Consumer Price Data," *Journal of Economic Perspectives*, 20, 171–192.

Diewert, E. W. (2012), "Consumer Price Statistics in the UK," Tech. rep., Office for National Statistics.

Duin, R. P. W. (1976), "On the Choice of Smoothing Parameters for Parzen Estimators of Probability Density Functions," *IEEE Transactions on Computers*, 25, 1175–1179.

Gagnon, E. (2009), "Price Setting during Low and High Inflation: Evidence from Mexico," *The Quarterly Journal of Economics*, 124, 1221–1263.

Huynh, K. P. and Jacho-Chávez, D. T. (2010), "Firm Size Distributions through the Lens of Functional Principal Components Analysis," *Journal of Applied Econometrics*, 25, 1211–1214.

Huynh, K. P., Jacho-Chávez, D. T., Kryvtsov, O., Shepotylo, O., and Vakhitov, V. (2016), "The Evolution of Firm-Level Distributions for Ukrainian Manufacturing Firms," *Journal of Comparative Economics*, 44, 148–162.

Huynh, K. P., Jacho-Chávez, D. T., Petrunia, R., and Voia, M. (2011), "Functional Principal Component Analysis of Density Families with Categorical and Continuous Data on Canadian Entrant Manufacturing Firms," *Journal of the American Statistical Association*, 106, 858–878.

Hyndman, R. J., Bashtannyk, D. M., and Grunwald, G. K. (1996), "Estimating and Visualizing Conditional Densities," *Journal of Computational and Graphical Statistics*, 5, 315–336.

Hyndman, R. J. and Khandakar, Y. (2008), "Automatic Time Series Forecasting: The Forecast Package for R," *Journal of Statistical Software*, 27, 1–22.

Kaplan, G., and Menzio, G. (2015), "The Morphology of Price Dispersion," *International Economic Review*, 56, 1165–1206.

Klenow, P. J. and Kryvtsov, O. (2008), "State-Dependent or Time-Dependent Pricing: Does It Matter for Recent U.S. Inflation?" *Quarterly Journal of Economics*, 123, 863–904.

Klenow, P. J. and Malin, B. (2010), "Microeconomic Evidence on Price-Setting," in *Handbook of Monetary Economics*, ed. Friedman, B. M. and Woodford, M., Elsevier, vol. 3, chap. 6, pp. 231–284.

Kneip, A. and Utikal, K. J. (2001), "Inference for Density Families Using Functional Principal Component Analysis," *Journal of the American Statistical Association*, 96, 519–542.

Kryvtsov, O. (2016), "Is There a Quality Bias in the Canadian CPI? Evidence from Microdata," *Canadian Journal of Economics*, 49, 1401–1424.

Kryvtsov, O. and Vincent, N. (2017), "The Cyclicality of Sales and Aggregate Price Flexibility," Mimeo, February.

Lach, S. and Tsiddon, D. (1992), "The Behavior of Prices and Inflation: An Empirical Analysis of Disaggregated Price Data," *Journal of Political Economy*, 100, 349–389.

Midrigan, V. (2011), "Menu Costs, Multi-Product Firms and Aggregate Fluctuations," *Econometrica*, 79, 1139–80.

Nakamura, E. and Steinsson, J. (2008), "Five Facts about Prices: A Reevaluation of Menu Cost Models," *Quarterly Journal of Economics*, 123, 1415–1464.

— (2013), "Price Rigidity: Microeconomic Evidence and Macroeconomic Implications," *Annual Review of Economics*, 213, 133–163.

ONS (2014), *Consumer Price Indices Technical Manual*, Office for National Statistics, Prices Division, Newport, United Kingdom.

Peterson, B. and Shi, S. (2004), "Money, Price Dispersion and Welfare," *Economic Theory*, 24, 907–932.

Piessens, R., de Doncker-Kapenga, E., Überhuber, C., and Kahaner, D. (1983), *Quadpack: A Subroutine Package for Automatic Integration*, Springer Series in Computational Mathematics, Berlin: Springer-Verlag.

Rosenblatt, M. (1956), "A Central Limit Theorem and a Strong Mixing Condition," *Proceedings of the National Academy of Sciences of the USA*, 42, 43–47.

Sheremirov, V. (2016), "Price Dispersion and Inflation: New Facts and Theoretical Implications," Mimeo, January.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall, 1st ed.

Thorarinsdottir, T. L., Gneiting, T., and Gissibl, N. (2013), "Using Proper Divergence Functions to Evaluate Climate Models," *SIAM/ASA Journal on Uncertainty Quantification*, 1, 522–534.

Tran, L. T. (1990), "Kernel Density Estimation on Random Fields," *Journal of Multivariate Analysis*, 34, 37–53.

Vavra, J. S. (2013), "Inflation Dynamics and Time-Varying Volatility: New Evidence and an Ss Interpretation," *The Quarterly Journal of Economics*, 129, 215–258.